

Однако в этой радужной перспективе есть черные пятна. Можно считать, что есть три основные трудности. Первая — небольшая клавиатура. Вторая — сравнительно небольшая площадь экрана. Третья — высокая стоимость передачи значительных массивов данных.

Нам представляется, что первая из указанных трудностей не очень существенна. Практика показывает, что у абонентов (особенно молодых) набор текстов на малой клавиатуре не вызывает особого дискомфорта. Набирается прямо-таки океан SMS сообщений.

Вторая трудность преодолима лишь частично. Производители телефонов стремятся увеличить площадь экрана, делая его открывающимся, однако, видимо, есть предел размеров для аппарата, обладающего свойством мобильности. Уменьшение символов тоже ограничено. Поэтому число знаков, помещающихся на экране, конечно, существенно меньше, чем на экране компьютера. Так что экран телефона всегда будет уступать по своим возможностям экрану стационарного компьютера.

И, наконец, высокая стоимость передачи данных не позволит пользователю спокойно “покопаться” в Интернете. Для компании надо будет сидеть за стационарный компьютер.

Вместо этого мобильный телефон в поисковом режиме должен обеспечить оперативное получение нужных сведений. Поэтому необходимо так организовать работу комплекса “мобильный телефон — сервер”, чтобы минимизировать количество информации в ответе на запрос при обеспечении высокой релевантности.

Оставляя пока в стороне требования к программным средствам обработки запроса, предположим, что в сформированном “конкретном запросе”, который поступил в поисковую систему, полно представлены все стороны информационной потребности пользователя. Если в дескрипторной системе с инвертированным поиском есть требуемый ответ (документ), то его название будет быстро найдено. Однако при этом будут найдены заодно названия и тех документов, которые не полностью удовлетворяют пользователя, или совсем ему не нужны, но содержат в себе те слова (дескрипторы), которые имеются в запросе. Это вследствие того, что система инвертированного поиска принципиально не может более точно оценить степень соответствия запросу каждого из документов найденного множества.

Поэтому сейчас в результате работы поисковой системы Интернета обычно выдается строка с указанием общего количества найденных документов, число сайтов, на которых они размещены, и постранично выдаются названия всех этих документов. Предлагается пользователю найти ответ на свой запрос “копаясь” в текстах документов из этого перечня. Т. е. путем последовательного просмотра найти тот самый документ, который его удовлетворяет.

В этой технологии дескрипторная поисковая система, конечно, сделала огромное дело, быстро выделив из огромного множества документов те, которые имеют отношение к запросу. Далее пользователь осуществляет прямой поиск в документах, используя, конечно, средства вызова текста, перемещения его по экрану, подсветку дескрипторов и так далее. Но если прямой поиск в текстах документов еще допустим (?) в стационарном варианте,

то никак не подходит для мобильного из-за высокой стоимости перекачки текстов, малой площади экрана, на котором даже названий документов поместится немного, и, наконец, из-за того, что абонент, вероятно, пожелает получить информацию за существенно меньшее время.

Поэтому имеющуюся технологию предлагается усовершенствовать путем замены прямого поиска, осуществляемого пользователем вручную, на автоматический прямой поиск. И выполнять его на сервере перед выдачей результата абоненту через канал связи. Если этот автоматический прямой поиск достаточно хорош (обеспечивает релевантность), то все найденные документы будут упорядочены по степени их соответствия запросу. В начале этого списка должен стоять тот документ, который, как мы предположили, содержит требуемый ответ. Можно провозгласить как бы лозунг: “Конкретный запрос — локальный ответ”.

Эту ситуацию образно можно описать так. Есть океан. Дескрипторная система инвертированного поиска из всей этой воды уместит наполнить ведро такой водой, которая приблизительно соответствует по своему составу тому, что требуется. Далее система прямого поиска по капле через пипетку (канал связи) выдает воду из ведра пользователю. Вначале капли самые лучшие, а затем (если падо) все хуже по степени соответствия запрошенному составу.

Как же организовать прямой поиск с оценкой степени соответствия заданному запросу?

Для этого не удастся использовать оценки документов “в целом”, так как информация, полученная при их включении в систему (перечни дескрипторов документов — их поисковые образы), уже использована в Интернете стандартным инвертированным поиском при их отборе. Придется более детально анализировать документы. Для этого, собственно, и нужен прямой поиск, который позволяет более детально оценить на степень соответствия запросу те фрагменты документа, которые можно выделить программно (используя формальные правила), и на основании множества оценок фрагментов оценить близость к запросу каждого из найденных документов. Что же является такими фрагментами?

По формальным правилам из текстов можно выделить большинство фраз и абзацев. В ряде случаев можно выделить и части фразы — перечисления, подпункты, когда текст разывается “красной строкой”. Бывает, что некоторые части текста вообще не структурированы как фразы, а выделение фрагментов в них производится только красной строкой. Например, так нередко выделяются заголовки и оглавления — весьма семантически важные фрагменты. Можно использовать еще и анализ выделений, скажем, шрифтом, цветом, которые, возможно, помогут семантически точно оценить вхождения в текст именных словосочетаний (терминов) и т. д.

Будем полагать, что стоит рассматривать три типа фрагментов: первый — выделение внутри фразы, второй — фраза и третий — абзац. Оценка каждого из фрагментов определяется числом слов запроса, присутствующих в этом фрагменте. При этом оценки “старших” фрагментов (абзацев) согласуются с оценками фрагментов, в них входящих

(фраз). А общая оценка текста определяется, во-первых, максимальным значением оценки фрагмента и, во-вторых, общим числом фрагментов, имеющих оценки хоть и меньшие, но близкие. Как и во всех задачах поиска оптимальности, нахождение единственной оценки требует назначения числовых значений для “весов” различных параметров, на нее влияющих. Возможно, что “продвинутым” абонентам можно будет предоставить возможность выбора подходящих для них значений весов параметров.

Если в сервере, на который поступил запрос с мобильного телефона, произведено считывание текстов всех найденных документов и их оценка, как описано выше, то в качестве ответа абоненту будет поступать список названий документов, упорядоченный по мере убывания оценок. Можно вызвать их тексты. По желанию абонента документ может быть выдан не с начала, а например, с наиболее хорошего фрагмента, или руководствуясь “кучностью” расположения слов запроса в тексте.

Упорядоченные документы это как раз те “капли”, которые получили оценки путем анализа всей совокупности капель воды, входящей в ведро, выбранное из океана.

Для удобства выдачи может быть использована все чаще применяемая в мобильных телефонах система автоматического синтеза речи. С ее помощью можно сообщить абоненту, сколько всего найдено документов по его запросу, сколько из них имеют высокие оценки, в каком количестве сайтов они размещены. Эта информация может выдаваться до окончания прямого поиска по всем найденным документам, а затем корректироваться. И сам упорядоченный список названий документов может выдаваться до окончания поиска, а затем корректироваться с извещением абонента голосом о появлении более хороших документов, чем тот, который им просматривается в настоящий момент.

Следует заметить, что изложенные требования к прямому поиску подробно описаны в статье А. Лукиной и Г. Миронова*. К этому надо добавить, что с момента выхода в свет этой статьи авторами отлажен комплекс программ (“Система 8.90”), осуществляющий прямой поиск и упорядочивающий документы так, как описано выше.

Теперь следует вернуться к программным средствам обработки запроса. Как получить “конкретный запрос”?

Надо заметить, что далеко не всегда пользователь, обращающийся к Интернету, сразу правильно формулирует свой запрос. Нередко требуется конкретизация, сужение запроса, или же, наоборот, его расширение, вплоть до развертывания его в множество запросов. Например, если пользователь намеревается приобрести некоторый товар, то его, как правило, интересует цена. Поэтому в запросе

это надо указать. Заметим, что для этого и других случаев, когда пользователя интересуют числовые значения параметров, следует использовать программы поиска и выделения фактографической информации, которые, вследствие выполнения прямого поиска на сервере, целесообразно иметь и выполнять на нем же. Примером же развертывания запроса в множество конкретных запросов может служить первоначальный запрос — желание пользователя получить конкретные сведения, например, о предполагаемом путешествии (туре) по маршруту: Австралия — остров Пасхи — Южная Америка. Следовало бы развернуть этот запрос в множество конкретных с заменой в каждом из них названия континента Южная Америка на страну, в него входящую. Будет серия запросов: Австралия — остров Пасхи — Чили; Австралия — остров Пасхи — Боливия; Австралия — остров Пасхи — Перу; .. Австралия — остров Пасхи — Аргентина. Для этого, понятно, должен использоваться географический классификатор. Целесообразно применения разного рода словарей и классификаторов для диалога с пользователем подробнее обсуждается в указанной статье. Здесь же надо отметить, что существенная часть этого диалога при задании запроса с мобильного телефона может осуществляться с помощью систем не только синтеза, но и анализа речи. Синтез уже прочно вошел в практику, а анализ (пословный) имеет удачные примеры использования.

До сих пор мы говорили об использовании справочной службы Интернет с мобильного телефона. Не меньшее значение может иметь ввод информации абонентом в Интернет. Надо сказать, что для ввода сравнительно коротких текстовых сообщений нет особых трудностей. С достаточно “продвинутых” аппаратов нетрудно вводить и видеoinформацию. Можно будет вводить и значительный текстовый (посимвольно кодированный) материал путем подключения к телефону устройств памяти. Однако надо особо отметить широко открывающиеся возможности ввода фактографической информации для систем типа электронной доски объявлений. Чтобы эти возможности не вызвали трудностей у абонента (теперь выступающего в роли корреспондента Интернет), целесообразно принять естественную модель представления фактографической информации, каковой являются объектно-характеристические таблицы.

Охарактеризованное программное обеспечение позволит сделать всемирную паутину хранилищем сведений о всех тех сторонах жизни великого множества людей, которые они посчитают возможным показать другим. И каждый сможет точно и быстро узнать все то, что ему интересно.

Материал принят к опубликованию 20.12.06.

*Лукина А., Миронов Г. Четыре программных средства, необходимых Интернету-2 // Информатизация и связь. 2005. № 1.